(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(51) Int. Cl.[6]
G06F 11/34

(11) 공개번호 특2000-0064410
(43) 공개일자 2000년11월06일

(21) 출원번호 10-1998-0704500
(22) 출원일자 1998년06월15일
번역문제출일자 1998년06월15일
(86) 국제출원번호 PCT/US1996/19810 (87) 국제공개번호
(86) 국제출원출원일자 1996년12월11일 (87) 국제공개일자
(81) 지정국 EP 유럽특허 : 오스트리아 벨기에 스위스 독일 덴마크 스페인 프랑스 영국 그리스 이탈리아 룩셈부르크 모나코 네덜란드 포르투칼 스웨덴
국내특허 : 아일랜드 오스트레일리아 브라질 캐나다 일본 대한민국 멕시코
(30) 우선권주장 8/573,127 1995년12월15일 미국(US)
(71) 출원인 마이렉스 코포레이션 콜린 그레이
미국 캘리포니아 94555 프레몬트 아덴우드 블러바드 34551
(72) 발명자 나가라즈 에스와드
미국 캘리포니아 94555 프레몬트 휭브렐 로드 33701
브하스카 에스오크
미국 캘리포니아 94555 프레몬트 허프만 테라스 4869
(74) 대리인 장용식

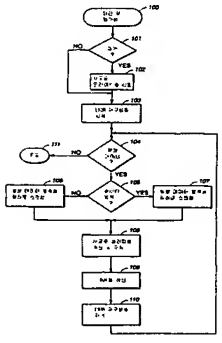
심사청구 : 없음

(54) 레이드시스템의불량데이터를관리하는방법및장치

요약

불량 디스크 데이터의 재구성동안 판독 에러가 발생할 때, 에러 블록에 따른 블록은 대응하는 불량 디스크 블록의 재구성을 가능하게 하지 않는다. 2개의 데이터 블록의 오용을 방지하기 위해, 불량 데이터 테이블[BDT](23)이 구성되며 블록의 어드레스 목록은 판독하며 그 블록은 재구성될 것이다. 표준 필러 블록(106)은 2개의 불량 블록으로 기록되며 새로운 패러티 블록이 생성된다. 메모리 어레이(200)에 대한 모든 액세스 요구의 어드레스는 BDT(109)와 비교되며, 리스트되지 않는다면, 그 액세스는 진행된다(200). 어드레스가 리스트되면, 에러 신호(204)는 복귀된다. 리스트된 기록 요구(203)에 있어서, 불량 블록 어드레스는 BDT(205)에서 삭제되고, 새로운 데이터가 블록으로 기록(206)되며 새로운 패러티 블록이 계산되어 저장된다.

대표도



명세서

기술분야

본 발명은 값싼 디스크의 여분 어레이(RAID)라고 불리는 하드 디스크의 불량 허용한계 어레이에 관한 것이다. 좀더 상세하게는, 본 발명은 미검출된 판독 에러를 포함하는 시스템이 아닌, 디스크 불량으로 인해 디스크의 교체 및 증설이 필요할 때 발견된 RAID 시스템에서 미검출된 판독 에러의 관리 및 수정에

관한 것이다.

발명의 요약

본 발명은, 어레이와 불량의 한 채널에서 디스크 기록 매체 에러, 및 다른 채널에서 디스크 드라이브의 재건때문에 같은 RAID 데이터 그룹내에 2개의 불량 데이터 블록이 존재할 때, 불요(不要) 데이터의 잠재적 생성을 회피하는 불량 데이터 관리 서브시스템을 제공한다. 본 발명의 불량 데이터 관리 서브시스템은:

a) 불량 디스크에서 데이터를 재구성하기 위해 사용되는 채널중 불량 데이터 블록을 검사하는 불량 드라이브를 재건하고, 불량이 발생하면, 그 블록 및 재건되고 있는 대응하는 디스크의 블록에 필러(FILLER) 코드를 기록함으로써 불량 데이터 블록을 스크럽(SCRUB)하기 위한 디스크 재건 루틴;

b) 필러 코드를 포함하는 모든 채널을 사용함으로써 불량 데이터 블록에 대응하여 새로운 패러티 블록을 계산;

c) 모든 불량 데이터 블록을 리스트하는 불량 데이터 테이블을 갱신; 및

d) 그 요구의 데이터 어드레스가 불량 데이터 테이블에 리스트되는지를 결정하기 위해 각 디스크 어레이 엑세스 요구를 검사하고, 리스트되지 않는다면 액세스가 진행 가능하도록 하고, 그렇지 않다면 액세스 요구가 기록용인지를 검사하며, 기록용이라면 기록을 가능하게 하며 불량 데이터 테이블의 블록 어드레스를 삭제하고 새로운 패러티 블록을 생성하며, 기록용이 아니라면 에러 신호를 생성하는 것을 포함한다.

본 발명은 상기 설명된 잠재적인 문제를 초래하지 않고 RAID 시스템에서 불량 데이터를 관리하는 수단을 제공한다.

바이트 스트라이프가 단일 어드레스 블록이기에, RAID-3 및 RAID-4 시스템이 블록길이를 제외하고 동일하다고 볼 수 있다. 이에 따라, 블록에 대한 다음에 따르는 모든 참조문은 다른 지시사항이 없다면 RAID-3 바이트 스트라이프를 포함하여 이해될 것이다.

또한, RAID 시스템은 다른 지시사항이 없다면 RAID-3, 4, 5 시스템을 나타내는 데 사용될 것이다.

개인용 컴퓨터를 위해 개발된 자기 디스크 기술에 기초한 RAID 시스템은 개선된 성능, 신뢰성, 전력 소비를 제공함으로써 고가의 디스크 메모리를 선별할 수 있는 주목할만한 대안을 제시한다. 소형 디스크의 제조자는 ANSI X3.131-1986 소형 컴퓨터 동기 인터페이스(SCSI)와 같은 더 높은 수준의 주변기기 인터페이스를 나타내는 표준화 노력으로 그런 성능을 제시할 수 있다. 이것은 데이터의 대규모 블록 전송을 위해 인터리브된 방식으로 구성된, 또는 소규모 전송 프로세싱을 위해 독립 병렬 액세스로 정렬된 값싼 디스크의 어레이의 발전을 초래했다.

대규모 디스크 어레이의 형성은 전기 기계 디바이스의 대규모 중복 사용으로 인해 발생되는 신뢰성 문제를 초래한다. 어레이의 평균 고장 시간(MTTF)은 어레이에 있어서 디스크 수와 함께 증가되는 것으로 판단되어 왔다(PATTERSON, D.A., GIBSON, G. 및 KATZ, R.H, 'A Case for Redundant Arrays of Inexpensice Disks(RAID)', 보고서 번호 UCB/CSD 87/391, 1987년 12월, 캘리포니아주 94720 버클리에 있는 캘리포니아 대학 컴퓨터과학과(EECS)).

도면의 간단한 설명

본 발명은 아래 주어진 상세한 설명 및 발명의 바람직한 실시예의 첨부된 도면으로부터 완전히 이해될 수 있으나, 특정 실시예에 따른 발명을 제한하도록 취급되선 안되며 설명 및 이해하기 위해서이다.

도 1은 디스크 어레이 컨트롤러를 포함하여 종래 기술 MxN 디스크 어레이의 블록도이다.

도 2는 도 1의 어레이에 기초한 종래 기술 논리 디스크 영역 구성을 도시한다.

도 3a는 종래 기술 RAID-1 시스템의 메모리 맵을 도시한다.

도 3b는 종래 기술 RAID-3 시스템의 메모리 맵을 도시한다.

도 3c는 종래 기술 RAID-4 시스템의 메모리 맵을 도시한다.

도 3d는 종래 기술 RAID-5 시스템의 메모리 맵을 도시한다.

도 4a, 4b 및 4c는 종래 기술에 따라, 각각 초기 RAID-1 메모리 맵, 이중 불량 메모리 RAID-1 메모리 맵 및 재건된 이중 불량 RAID-1 메모리 맵을 도시한다.

도 5a, 5b 및 5c는 종래 기술에 따라, 각각 초기 RAID-5 메모리 맵, 이중 불량 RAID-5 메모리 맵 및 재건된 이중 불량 RAID-1 메모리 맵을 도시한다.

도 6은 본 발명의 디스크 재건 및 재구성하는 흐름도이다.

도 7은 어레이를 액세스하는 본 발명의 방법을 도시하는 흐름도이다.

도 8은 RAID 불량 데이터 관리 시스템을 위한 본 발명의 하드웨어 구성을 도시한다.

발명의 상세한 설명

도 6은, 다른 채널에서 대응하는 데이터 블록을 재구성하는동안 디스크 판독 불량이 RAID 시스템의 채널에서 발생하였기에, 불요(不要) 데이터가 생성되는 문제를 회피하는 데 사용되는 디스크 재건 및 재구성 방법(100)을 도시하는 흐름도이다. 데이터 블록의 재구성 또는 디스크 드라이브의 재건은 공통 패러티 블록을 공유하는 같은 그룹에 속하는 잔류 블록의 대응하는 비트를 비트 대 비트로 EXOR논리비교함으로

써 완성된다.

디스크 재건 및 재구성 방법(100)은 재구성을 EXOR 방법으로 시작하는 스텝(101)에서 시작된다. 재구성 동안, 스텝(102)은 재구성을 위해 필요한 블록중 하나가 불량이라는 지시를 검사한다. 불량 블록이 검출되지 않고 재구성이 완료되면, 프로세스는 스텝(111)에서 종료된다. 그렇지 않다면, 스텝(105)은 재구성을 위해 필요한 불량 데이터 블록이 패러티 블록인지를 검사한다. 패러티 블록이라면, 스텝(104)은 재구성되는 블록에 필러 블록을 기록하고 스텝(105)은 그 데이터 블록이 회복불능(불량)이라고 기록함으로써 불량 데이터 테이블(BDT)을 갱신한다. (필러 블록은 다른 편리한 영숫자(英數字) 코드를 포함할 수 있다) 스텝(103)의 불량 데이터 블록이 패러티 블록이 아니라면, 스텝(106)은 불량 데이터 블록을 교체하기 위해 필러 블록을 기록함으로써 불량 데이터 블록을 스크럽하며, 스텝(107)에서, 재구성되었을 대응하는 블록에 필러 블록을 또한 기록한다. 스텝(107)은 스텝(103)의 블록 및 재구성되고 있는 채널의 대응하는 블록이 불량임을 기록함으로써 BDT를 갱신한다. BDT의 각 엔트리는 인공 데이터 및 회복불능 데이터를 포함함으로써 블록을 식별한다. 스텝(109)은 모든 데이터 채널 블록(필러 블록을 포함하여)을 EXOR논리비교함으로써 새로운 패러티 데이터를 계산하고, 스텝(100)은 새로운 패러티 블록을 기록하며 스탭(100)은 스텝(102)으로 복귀하여 재구성 프로세스를 재개한다.

도 6의 절차는

(1) 데이터 블록중 하나에서 재구성동안 판독 에러가 발생한다면, 모든 잔류하는 양호한 데이터 블록은 회복가능하고,

(2) 필러 블록의 인공 데이터는 실 데이터와 혼돈될 수 없다는 것을 확실히 해준다.

스텝(109)은 디스크 컨트롤러(12)에 위치하는 불량 데이터 테이블, 즉 BDT를 갱신하고, 각 디스크의 예약 구간에도 위치할 수 있다. BDT는 모든 불량 데이터 블록(필러 블록)의 리스트를 포함하여, 디스크 어레이에 대한 어떠한 판독 요구라도 목표 어드레스를 BDT 어드레스에 리스트된 어드레스와 먼저 비교하고, 어드레스가 리스트되어 있다면 적절한 에러 신호를 요구 에이전트로 복귀시킨다. BDT가 갱신된 후, 재구성 프로세스는 스텝(110)에서 계속되며, 스텝(104)의 감시는 완료될때까지 계속된다. 추가 불량 데이터 블록이 스텝(104)에서 발견된다면, 스텝(105-110)은 필요한만큼 반복된다.

도 7의 흐름도에 도시된 바와 같이, 어레이 액세스(200)를 검사하는 비교기를 동작하기위한 방법은 RAID 메모리 시스템과 함께 BDT 사용을 관리하는데 이용된다. 그 방법은 액세스가 요구될때마다 호출되며, 액세스 요구가 어드레스가 BDT에 리스트되어 있는 불량 데이터 블록에 대한 것인지를 검사함으로써 스텝(201)에서 시작되고, 그렇지 않다면 정상 동작 모드의 액세스로 진행한다. 어드레스가 BDT에 리스트된다면, 스텝(203)은 액세스가 기록 동작용임을 결정하고, 그렇지 않다면, 스텝(204)에서 에러 플래그가 생성된다. 에러 플래그는 판독 요구는 회복불능 데이터 블록에 대한 것임을 호스트 시스템에게 알려준다. 요구가 기록 액세스용이라면, 스텝(205)은 BDT로부터 블록 어드레스를 삭제하고 스텝(206)에서 기록 회로가 새로운 데이터를 블록 어드레스에 기록하도록 허용한다. 스텝(207)은 새로운 블록 패러티를 계산하고 이것을 액세스된 데이터 그룹과 연관된 대응하는 패러티 블록에 기록하기 위한 것이다.

도 8은 도 6 및 7에 개설된 방법을 이용하는 본 발명의 RAID 시스템의 아키텍쳐를 도시하는 블록도이다. 특히, 상기 설명된 종류의 불량 데이터 관리를 갖춘RAID 시스템(20)은 SCSI 버스(21)로부터 이것의 컨트롤러(22)를 통해 인터페이스된다. (SCSI는 미국 규격 X3T9.2/86-109 소형 컴퓨터 시스템 인터페이스-2에 설명된 것처럼 잘 알려진 업계 표준 버스를 나타낸다.) 버스(21)는 RAID 시스템(20)을 호스트 컴퓨터에 연결한다. 디스크 어레이 컨트롤러(22)는, 메모리 액세스와 연관된 어떤 논리 어드레스를 대응하는 물리 어드레스에 사상하고, 데이터 트래픽을 관리하고, 각 디스크 드라이브의 동작을 제어하고, 상태 정보를 호스트에 제공하기 위해 필요한 논리를 제공한다. BDT(23)는 컨트롤러(22)에 커플되어 도시되지만 집적 칩 디스크 어레이 컨트롤러의 필수 부분이 될 수 있다. 또한, 액세스 어드레스가 디스크 컨트롤러(22)의 어레이 레벨보다는 디스크 레벨에서 검사될 수 있도록 각 디스크의 불량 데이터 정보가 선택적으로 저장되는 로컬 BDT(24)를 갖는 어레이의 각 디스크(11)가 도시된다.

기술에 숙련된 이들에게는 이해될 것이며, 상기 설명한 방법 및 장치의 많은 변화는 본 발명의 사상 및 범위를 벗어나지 않고 숙련된 개업자에 의해 발생할 수 있으며, 다음에 따르는 청구항으로만 제한되어야 한다.

(57) 청구의 범위

청구항 1

메모리 시스템에 대한 액세스를 제어하고, 메모리 불량을 검출하며, 메모리 에러를 검출하고, 단일 채널을 정정하기 위한 메모리 어레이 컨트롤러를 갖는 다중채널 메모리 시스템의 불량 데이터를 관리하는 장치에 있어서, 각 채널은 적어도 하나의 메모리 모듈을 구비하며, 각 모듈은 불량 및 판독 에러 검출 수단을 구비하고, 각 메모리 모듈은 불량에 대해 독립적으로 교체될 수 있으며, 메모리 컨트롤러에 액세스 가능한 상기 장치는:

a) 하나의 채널이상에서 동시에 발생하는 불량에 의한 회복불능 데이터의 어드레스를 저장하는 불량 데이터 테이블(BDT);

b) 불량 데이터 블록을 교체 및 회복불능 데이터를 나타내는 불량 데이터 블록의 어드레스를 저장하는 BDT에 기록하기 위한 불량 블록 위치에 필러 데이터를 기록하는 기록 회로; 및

c) BDT에 저장된 어드레스에 대한 메모리 액세스 요구를 검출하고, 액세스 요구가 판독 요구라면 회복불능 데이터 에러 신호를 호스트 시스템에 복귀시키고, 액세스 요구가 기록 요구라면 BDT 리스트된 어드레스에 기록하고 BDT로부터 리스트된 어드레스를 삭제가능하게 하는 검출 회로를 포함하는 것을 특징으로 하는 불량 데이터 관리 장치.

청구항 2

제 1 항에 있어서, 다중채널 메모리 시스템이 RAID-3 메모리 어레이인 것을 특징으로 하는 불량 데이터 관리 장치.

청구항 3

제 1 항에 있어서, 다중채널 메모리 시스템이 RAID-4 메모리 어레이인 것을 특징으로 하는 불량 데이터 관리 장치.

청구항 4

제 1 항에 있어서, 다중채널 메모리 시스템이 RAID-5 메모리 어레이인 것을 특징으로 하는 불량 데이터 관리 장치.

청구항 5

제 1 항에 있어서, BDT가 메모리 어레이 컨트롤러 부분인 것을 특징으로 하는 불량 데이터 관리 장치.

청구항 6

제 1 항에 있어서, BDT가 각 메모리 모듈의 예약 구간에 저장되는 것을 특징으로 하는 불량 데이터 관리 장치.

청구항 7

제 1 항에 있어서, BDT가 메모리 어레이 컨트롤러의 메모리 및 각 메모리 모듈의 예약 구간에 저장되는 것을 특징으로 하는 불량 데이터 관리 장치.

청구항 8

메모리 에러 검출 및 단일 채널 정정 수단을 갖는 다중채널 메모리 시스템의 불량 데이터를 관리하는 방법에 있어서, 각 채널은 불량 및 판독 에러를 검출할 수 있는 적어도 하나의 메모리 모듈을 구비하며, 각 메모리 모듈은 불량에 대해 독립적으로 교체될 수 있고, 상기 방법은:

a) 소정의 채널의 블록에서 판독 에러를 검출하는 단계;

b) 판독 에러가 메모리 모듈로 인한 것이라면 불량이 나타난 메모리 모듈을 교체하며, 그렇지 않다면 소정의 채널에서 데이터를 재구성하기 위한 단일 채널 정정 수단을 사용하는 것을 결정하는 단계;

c) 데이터 재구성을 방해하는 제 2 판독 에러에 대한 다른 모든 채널을 감시하고, 그러한 제 2 판독 에러가 발견되고 제 1 또는 제 2 블록 어느 것도 패러티 블록이 아니라면, 제 2 블록 어드레스로 진입하고, 제 1 블록의 어드레스를 불량 데이터 테이블(BDT)과 대응시키며, 제 1 및 제 2 블록에 필러 블록을 기록하고, 제 1 및 제 2 블록을 대응하는 모든 데이터 블록과 함께 사용하여 대응하는 패러티 블록을 계산 및 교체하며, 그렇지 않다면 제 1 블록 및 제 2 블록중 어느 하나가 데이터 블록인 곳에 필러 블록을 기록하고, 대응하는 모든 데이터 블록과 함께 필러 블록을 사용하여 대응하는 패러티 블록을 계산 및 기록하며, 그 후 소정의 채널에서 완료될때까지 데이터를 재구성하는 단계; 및

d) BDT 리스트된 어드레스와 요구된 어드레스를 비교함으로써 모든 메모리 액세스 요구를 감시하며, 그 어드레스가 BDT에 리스트되어 있지 않다면 액세스 요구가 진행되도록 하며, 그렇지 않다면 요구가 기록 요구인지를 검사하고, 기록 요구가 아니라면 회복불능 데이터 에러 신호를 복귀시키며, 기록 요구라면 어드레스 기록되어 있는 필러 블록이 유효 데이터 블록에 의해 교체되고 그 후 대응하는 새로운 패러티 데이터 블록으로 진입하고 계산하기 위해 기록 요구가 진행되도록 하는 단계를 포함하는 것을 특징으로 하는 불량 데이터 관리 방법.

청구항 9

제 8 항에 있어서, 다중채널 메모리 시스템이 RAID-3 메모리인 것을 특징으로 하는 불량 데이터 관리 방법.

청구항 10

제 8 항에 있어서, 다중채널 메모리 시스템이 RAID-4 메모리인 것을 특징으로 하는 불량 데이터 관리 방법.

청구항 11

제 8 항에 있어서, 다중채널 메모리 시스템이 RAID-5 메모리인 것을 특징으로 하는 불량 데이터 관리 방법.

청구항 12

다중 불량을 처리하는 불량 데이터 관리 시스템을 갖춘 다중채널 메모리 어레이 시스템에 있어서,

a) 각 채널이 적어도 하나의 메모리 모듈을 가지고, 각 메모리 모듈이 불량에 대해 독립적으로 교체될 수 있으며, 각 모듈이 불량 및 판독 에러 검출기를 갖는 메모리 어레이; 및

b) 메모리 어레이에 대한 액세스를 제어, 메모리 어레이 불량 검출과 단일 채널 정정, 및 메모리 버스를 통해 메모리 상태를 통신하기 위한, 메모리 버스에 메모리 어레이를 연결하고,

i) 하나 이상의 채널에서 동시에 발생하는 불량으로 인한 회복불능 데이터 블록의 어드레스를 저장하는

불량 데이터 테이블(BDT);

ii) 회복불능 데이터 블록에 대응하는 불량 데이터 블록에 필러 데이터를 기록하고 및 BDT에 회복불능 데이터 블록 어드레스를 저장하는 기록 회로; 및

iii) 메모리 액세스 요구 어드레스를 BDT에 저장된 어드레스와 비교함으로써 불량 데이터 블록에 대한 메모리 액세스 요구를 검출하고, 액세스 요구가 판독 요구이고 그 어드레스가 BDT에 저장되어 있다면 회복불능 에러 신호를 요구하는 에이전트에 복귀시키고, 요구가 기록 액세스에 대한 것이라면 액세스 요구가 진행되도록 하는 비교기를 더 구비한 메모리 어레이 컨트롤러를 포함하는 것을 특징으로 하는 다중채널 메모리 어레이 시스템.

청구항 13

제 12 항에 있어서, 각 메모리 모듈이 BDT를 저장하기 위한 예약된 메모리 영역을 갖는 것을 특징으로 하는 다중채널 메모리 어레이 시스템.

청구항 14

제 12 항에 있어서, 다중채널 메모리 어레이 시스템이 RAID-3 메모리 시스템인 것을 특징으로 하는 다중채널 메모리 어레이 시스템.

청구항 15

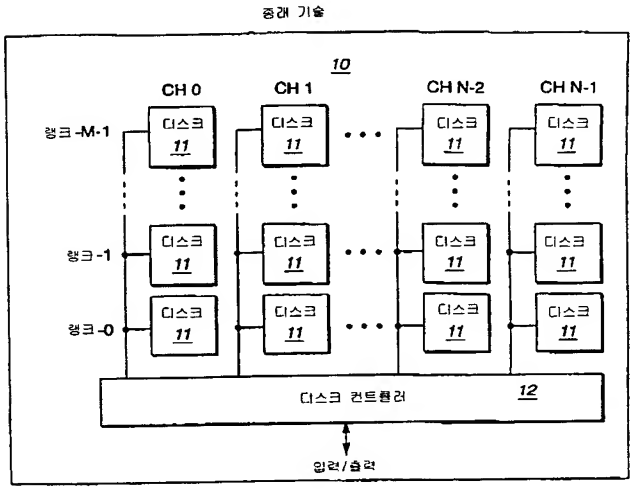
제 12 항에 있어서, 다중채널 메모리 어레이 시스템이 RAID-4 메모리 시스템인 것을 특징으로 하는 다중채널 메모리 어레이 시스템.

청구항 16

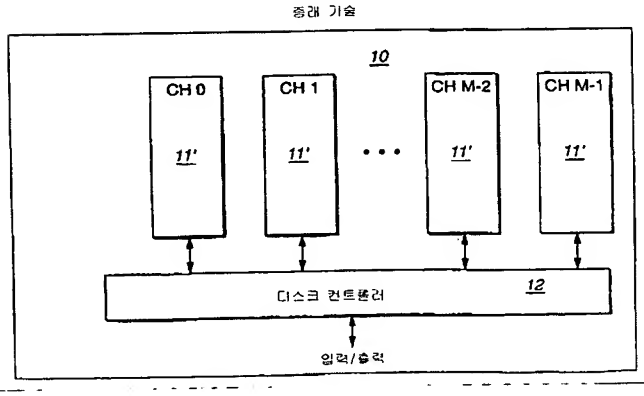
제 12 항에 있어서, 다중채널 메모리 어레이 시스템이 RAID-5 메모리 시스템인 것을 특징으로 하는 다중채널 메모리 어레이 시스템.

도면

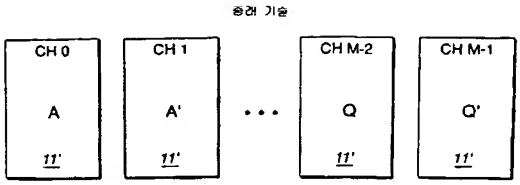
도면1



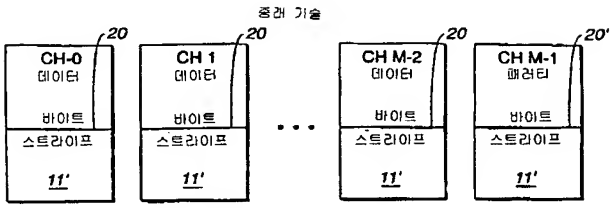
도면2



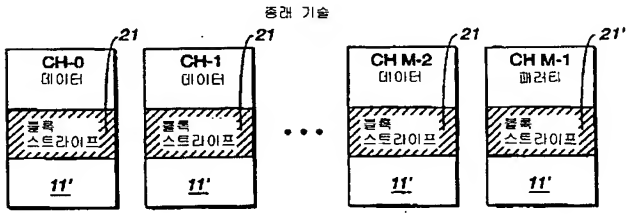
도면3a



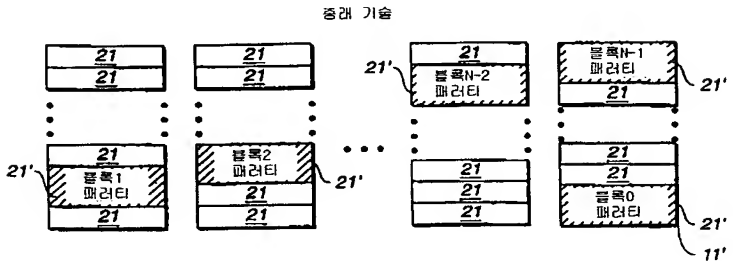
도면3b



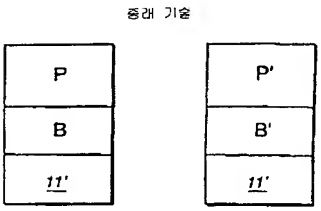
도면3c



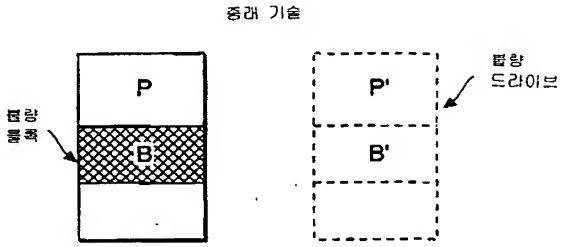
도면3d



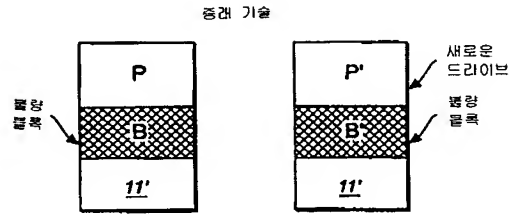
도면4a



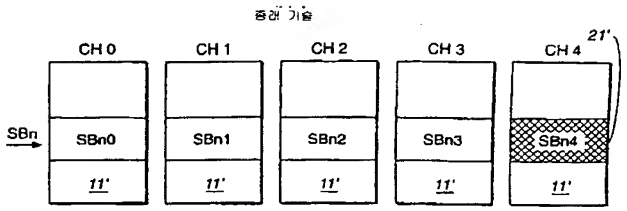
도면4b



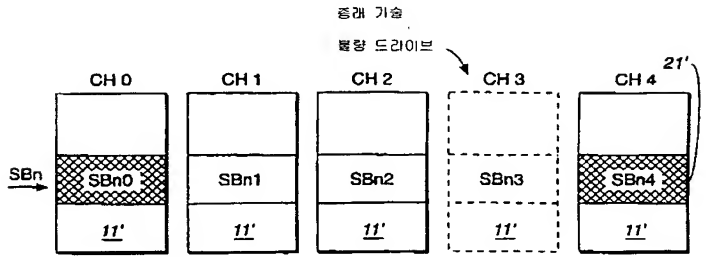
도면4c



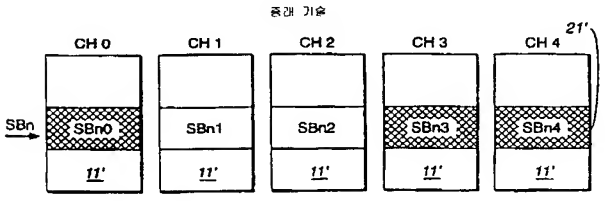
도면5a



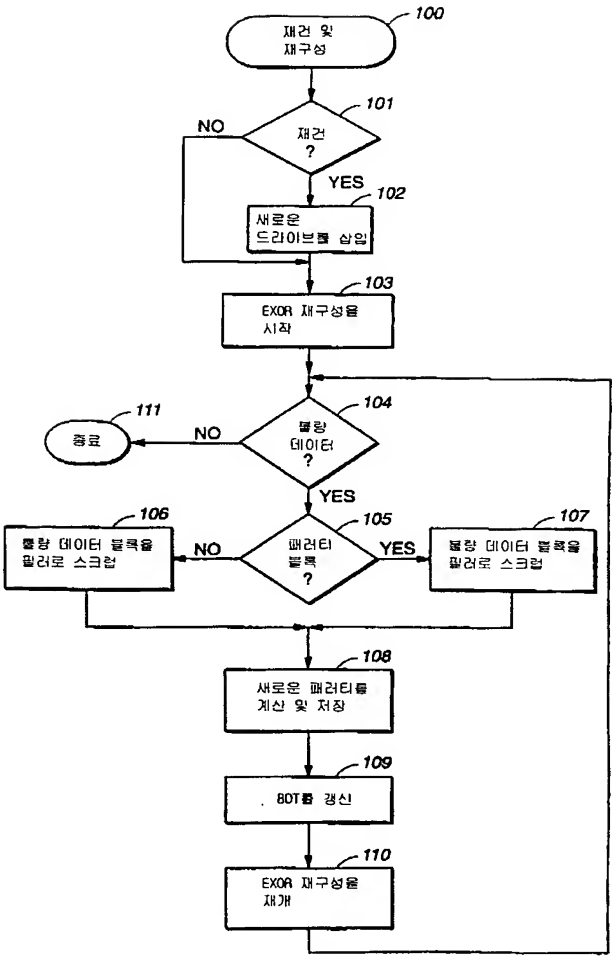
도면5b



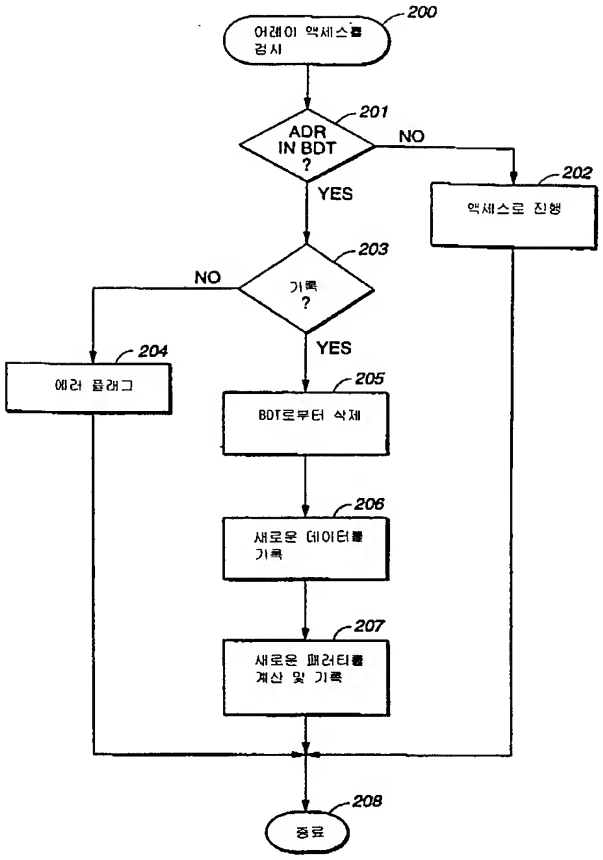
도면5c



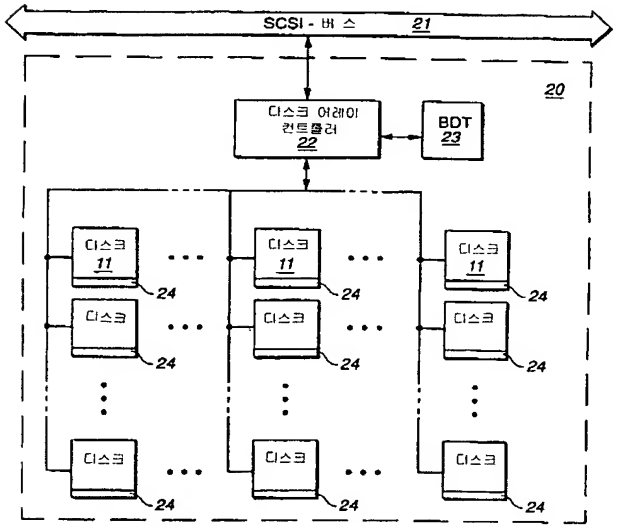
도면6


재건 및 재구성
100
재건 ?
101
NO
YES
새로운 드라이브를 삽입
102
EXOR 재구성을 시작
103
불량 데이터 ?
104
NO
종료
111
YES
패러티 블록 ?
105
NO
YES
불량 데이터 블록을 팔러로 스크랩
106
불량 데이터 블록을 팔러로 스크랩
107
새로운 패러티를 계산 및 저장
108
BDT를 갱신
109
EXOR 재구성을 재개
110

도면7


200
어레이 액세스를 검사
201
ADR IN BDT ?
NO
YES
202
액세스로 진행
203
기록 ?
NO
YES
204
에러 플래그
205
BDT로부터 삭제
206
새로운 데이터를 기록
207
새로운 패러티를 계산 및 기록
208
종료

도면8


SCSI - 버 스 21
20
디스크 어레이 컨트롤러 22
BDT 23
디스크 11
24